SOME PROBLEMS IN THE ANALYSIS OF STOCHASTIC DOMINANCE

William E. Stein

Department of Business Analysis, Texas A&M University
College Station, Texas 77843 USA

Roger C. Pfaffenberger

M. J. Neeley School of Business, Texas Christian University
Fort Worth, Texas 76129 USA

Abstract. The use of stochastic dominance has become common in finance and economics. As a theoretical device it is used to define a preference relation on a set of decision alternatives, thereby reducing the number of these alternatives which must be considered further by the decision-maker. However, in practice, data must be collected to estimate probability distributions. The paper discusses the errors which may result and the computation of their probabilities. The connection with statistical hypothesis testing is discussed.

Keywords. Kolmogorov-Smirnov test; hypothesis testing; sampling errors.

INTRODUCTION

When a decision-maker (DM) is faced with a choice between several risky alternatives it is common to approach the problem using the expected utility model. With this approach the decision-maker treats the payoff under each decision alternative as a random variable. Given his utility function, the DM can compute the expected utility for each alternative and choose the one with the greatest value. Even if the utility function is not known exactly but only some of its properties are known then it may still be possible to give a partial solution to the decision problem. This would result in the elimination of some inferior decision alternatives. The DM then need only concentrate on making a choice from the remaining ones. Stochastic dominance (SD) is a method of comparison of random variables which permits this elimination of decision alternatives.

If the distribution functions (c.d.f.) of the random variables are not known then stochastic dominance techniques cannot be applied. In that case, historical data is examined to give partial information about the c.d.f.'s. Of course, using this (imperfect) information we may infer SD holds when it doesn't (in the "population") and conversely. In this paper, we examine this problem in detail. In particular, how common would these errors be and what could be done to improve the situation? In the next section we give precise definitions of all terms and concepts and put the problem in mathematical form. Following that, we examine what has been accomplished in the finance literature. Finally, we discuss the problem from the view of statistical hypothesis testing.

BASIC CONCEPTS

Let random variables X, Y represent the (stochastic) return on each of two possible investment portfolios over a specified horizon. Then for a given utility function u, X is preferred to Y if:

$$E[u(X)] \geq E[u(Y)] \qquad (1)$$

Even if only general properties of u are known, we may still be able to determine preference. Let F, G be population c.d.f.'s (assumed continuous) corresponding to the random variables X, Y respectively. Then we may say that F dominates G (written FDG) by:

(A) First Order Stochastic Dominance (FSD) if $F(x) \leq G(x)$ for all x. FDG by FSD if and only if (1) is true for all non-decreasing utility functions u.

(B) Second Order Stochastic Dominance (SSD) if

$$\int_{-\infty}^{x} F(z)dz \leq \int_{-\infty}^{x} G(z)dz, \qquad \text{for all } x.$$

FDG by SSD if and only (1) is true for all non-decreasing concave utility functions.

FSD is appropriate for any DM since it only assumes that more (money) is preferred to less. SSD, on the other hand, corresponds to a DM who is risk-averse. In either case, if FDG then the investment corresponding to G need not be considered further since it is dominated by the investment corresponding to F. (Note that if F=G then FDG and GDF as opposed to the usual convention of no dominance in this case. Of course, in this case, whether we eliminate F or G or neither is irrelevant.)

Now assume F, G are unknown. To check for FSD or SSD we obtain a sample of observations from these distributions. Let F_m, G_n represent the empirical c.d.f.'s of F, G based on respective sample sizes of m and n. Historically the procedure has been to replace F, G in the definitions with F_m, G_n to test if dominance exists. For FSD, this reduces to checking if F_m is always above (or always below) G_n or if it crosses. Consequently, we shall refer to this approach as the "crossing algorithm." This method can be traced to Levy and Hanoch (1970) who used it for illustrative purposes only. However, later researchers applied this method without fully appreciating the potential for making an error. That is, given F, G and sample sizes m, n, there are three mutually exclusive and exhaustive events that may occur:

$$\text{(A) } F_m D G_n \qquad \text{(B) } G_n D F_m \qquad \text{(C) Crossing} \qquad (2)$$

MM8:1987-U

where "crossing" indicates no dominance in the sample. We certainly would like some idea of the probabilities of these events (assuming a given functional form for F and G).

SIMULATION RESULTS

We now give a brief discussion of the estimation of the probabilities of the events in (2) obtained in the finance literature by simulation. For details, see Dickinson (1974), Johnson and Burgess (1975), Knoll and Levy (1980), and Pope and Ziemer (1984).

Suppose FDG with F, G specified distributions. When F_n, G_n are obtained there are three possible conclusions given by (2). Here A is the correct conclusion while B, C are incorrect. Using computer simulation the probability of errors B and C can be estimated.

For example, assume F is normal with mean 1.1 and variance 1.0 while G is normal with mean 1.0 and variance 1.0. Then Pope and Ziemer (1984) chose n=m equal to various sample sizes from 5 through 100 and performed the simulation 250 times for SSD. Between 33% and 46% of the samples gave the correct conclusion (A), with no clear dependence on sample size. The reverse dominance (B) was found in 10% and 22% of the samples, generally decreasing with sample size. This means that no dominance (C) was found in 23% to 54% of the samples, increasing with sample size. Thus, increasing sample size does not always decrease the probability of an incorrect response. This is consistent with the findings of Kroll and Levy (1980) for FSD and SSD. Certainly, the high error probabilities reported in these simulation studies should be cause for serious concern.

HYPOTHESIS TESTING

In this section we will examine the FSD crossing algorithm from a statistical viewpoint. The crossing algorithm can be viewed as a hypothesis test related to the Kolmogorov-Smirnov (K-S) test. Let $F_n(x)$, $G_n(x)$ be the empirical c.d.f.'s considered as functions of x. To show the connection with the K-S test let

$$D^+_{nn} = \sup_x[F_n(x) - G_n(x)] \geq 0 \tag{3a}$$

$$D^-_{nn} = \sup_x[G_n(x) - F_n(x)] \geq 0 \tag{3b}$$

Then $P(F_nDG_n) = P(D^+_{nn}=0)$ (4a)

$$P(G_nDF_n) = P(D^-_{nn}=0) \tag{4b}$$

$$P(\text{Crossing}) = P(D^+_{nn}>0 \text{ and } D^-_{nn}>0) \tag{4c}$$

This is analogous to the two-sample K-S test of H_0:F=G but without the usual "confidence band." In other words, if $F_n(x)$ is ever greater than $G_n(x)$, even if only by a very small amount, then we will never be able to conclude FDG. It is therefore not surprising that the crossing algorithm is known to be a very sensitive test. We now consider two cases: whether F and G are equal or not.

The Case F=G

Stein, Pfaffenberger, and Kumar (1983) have shown the following (for equal sample sizes m=n):

$$P(F_nDG_n) = 1/(n+1) \tag{5a}$$

$$P(G_nDF_n) = 1/(n+1) \tag{5b}$$

$$P(\text{Crossing}) = 1 - 2/(n+1) \tag{5c}$$

(Solutions also have been obtained for values of m and n, m≠n). If H_0:F=G is taken to mean no dominance than P(error) = $P(F_nDG_n \text{ or } G_nDF_n)$ = 2/(n+1). As can be seen, under H_0 the error probability is independent of the common value of F and G.

The Case F≠G

We now take F≠G as the null hypothesis. While a more realistic situation than assuming F=G, this introduces considerable mathematical complications. This is equivalent to computing the power of the K-S test, which is known to be a difficult problem.

Steck (1974) has computed the power of a K-S test for c.d.f.'s F and G, with $F=G^k$ where k>0 is any positive real number. We can transform this problem to F(x)=x on (0,1) and $G(x) = x^{1/k}$ so that the error probabilities do not depend on the form of F and G except through the value k. Steck's method expresses $P(F_nDG_n)$ and $P(G_nDF_n)$ as a determinant. See Stein and Pfaffenberger (1982) for details. For arbitrary F and G no such solution is possible. In this case simulation is the only way to proceed, as in Kroll and Levy (1980).

WHITMORE'S TEST

As was mentioned in the previous section, the FSD crossing algorithm is a special case of the K-S test. One idea is to try to use a K-S test with a better choice of the confidence band. Whitmore (1978) suggested a 3-decision version of the K-S test. We present it here as a 4-decision problem. The test statistics are given in (3).

Choose a positive number d. The decision rule is:

If $D^-_{nn} > d$ and $D^+_{nn} > d$, decide no dominance.

If $D^-_{nn} > d$ and $D^+_{nn} < d$, decide FDG.

If $D^-_{nn} < d$ and $D^+_{nn} > d$, decide GDF.

Otherwise, no decision.

The last of the above cases corresponds to a failure to reject the null hypothesis H_0:F=G. This will occur if F_n and G_n are close over the entire range.

Whitmore suggests choosing d so that

$$P(\text{decide GDF or FDG} \mid F=G) = \alpha \tag{6}$$

where α is the desired Type I error probability. This can be done no matter what the common value of F and G is. The value of d was chosen in this manner since it was claimed that any non-dominant F, G would be less likely to lead to a conclusion of dominance than F=G. That is,

$$P(\text{decide GDF or FDG} \mid \text{any non-dominant } F,G) \leq \alpha \tag{7}$$

If true, this means that we can take H_0 to be "no dominance" and control the error probability by using the case F=G. However, this is false. Consider the case where a pair of single-crossing distributions with F exceeding G by only a small amount while, on another region, G exceeds F by a large amount. In fact, we can make F barely exceed G on one region so that D^+_{nn} will be small (<d) almost surely. For a fixed m=n, and α, this will lead to a FDG decision with a probability approaching 1.0 which contradicts (7). This means the test is biased (Massey, 1950).

An additional complication is that D^+, D^- are highly negatively correlated and consequently the rectangular critical regions are inappropriate. If (6) is used, it will be essentially impossible to reach the "no dominance" conclusion.

There are no known tests for FSD that are (a) distribution-free; (b) require only moderate sample sizes; and (c) are unbiased and consistent. In fact, it is not even clear what the null hypothesis should be. H_0: F=G seems reasonable but we cannot conclude (7) is true as we would hope. H_0: "no dominance" might be more reasonable but it is not clear how to handle this situation when computing the distribution of the test statistic under H_0.

One promising approach is that of Franck (1984) who developed a test of H_0: $F \geq G$ versus H_1: $F < G$. This test satisfies (a) and (c) of the above conditions but still requires large sample sizes.

HIGHER ORDER DOMINANCE

When we move to consideration of the SSD crossing algorithm, the situation is much more complex. First of all, even assuming that H_0: F=G will not produce error probabilities independently of F=G.

Instead, Whitmore (1978) suggested another test for SSD, an integrated version of the K-S test using the test statistic:

$$\sup_{x} \int_{-\infty}^{x} [F_m(z) - G_n(z)]dz \tag{8}$$

Unfortunately, this statistic is not distribution-free if F=G. If F=G then we can make this statistic distribution-free by replacing dz by dF. However, the statistic will no longer provide a reasonable test of SSD.

If we implement Markowitz's Mean-Variance rule by replacing the unknown population mean and variance by the sample mean variance then the estimation risk is dependent upon the population distribution. Consequently, it can be very difficult to draw general conclusions from simulation results (see Kroll & Levy (1980), Johnson & Burgess (1975), Frankfurter et all. (1971), Pope and Ziemer (1984)). However, since these studies all use a normal population the M-V rule is equivalent to SSD.

Recent work by Deshpande and Singh (1985) attacked the one-sample SSD problem. That is, assume F_0 is a known c.d.f.. To test H_0: F=G against the alternative that F dominates F_0 (SSD) they used an integrated analog of (8). This also is not distribution free under the null hypothesis. Further work is needed to extend this to the two-sample case.

CONCLUSIONS

The results of our work indicate that there is no acceptable way to use data to test for FSD. The situation for SSD is even worse, due to the mathematical difficulty of working with integrated c.d.f.'s. Since it may be expected that any distribution-free test for SSD that may be developed would perform more poorly than tests based on a known distribution, the outlook for a reasonable statistical analysis of the SSD problem is not bright. Extending the work of Franck (1984) may provide an acceptable approach to the FSD problem, but a solution for SSD seems elusive.

REFERENCES

Deshpande, J. V., and H. Singh, (1985). "Testing for second order stochastic dominance," Communications in Statistics - Theory and Methods, 14, 887-893.

Dickinson, J. P., (1974). "The reliability of estimation procedures in portfolio analysis,"Journal of Financial and Quantitative Analysis, 9, 447-462.

Durbin, J., (1973). Distribution Theory for Tests Based on the Sample Distribution Function, Society for Industrial and Applied Mathematics.

Franck, W. E., (1984). "A likelihood ratio test for stochastic ordering," Journal of American Statistical Association, 79, 686-691.

Frankfurter, G. M., H. E. Phillips and J. P. Seagle, (1971). "Portfolio selection: the effects of uncertain means, variances, and covariances," Journal of Financial and Quantitative Analysis, 6, 1251-1262.

Johnson, K. H. and R. C. Burgess, (1975). "The effects of sample sizes on the accuracy of EV and SSD efficiency criteria," Journal of Financial and Quantitative Analysis, 10, 813-820.

Kroll, Y. and H. Levy, (1980). "Sampling errors and portfolio efficient analysis," Journal of Financial and Quantitative Analysis, 15, 655-688.

Levy, H. and G. Hanoch, (1970). "Relative effectiveness of efficiency criteria for portfolio selection," Journal of Financial and Quantitative Analysis, 5, 63-76.

Massey, F. J., (1950). "A note on the power of a non-parametric test," Annals of Mathematical Statistics, 21, 440-443.

Pope, R. D., and R. F. Ziemer, (1984). "Stochastic efficiency, normality, and sampling errors in agricultural risk analysis," American Journal of Agricultural Economics, 66, 31-40.

Steck, G. P., (1974). "A new formula for $P(R_i < b_i, 1 \leq i \leq m | m,n,F=G^k)$," Annals of Probability, 2, 155-160.

Stein, W. E., R. Pfaffenberger, and P. C. Kumar, (1983). "Estimation risk in first order stochastic dominance: A note," Journal of Financial and Quantitative Analysis, 18, 471-476.

Stein, W. E., and R. C. Pfaffenberger, (1982). "An examination of tests for stochastic dominance," M.J. Neeley School of Business, Texas Christian University, Working Paper #3, 1982-83 Series.

Whitmore, G. A., (1978). "Statistical tests for stochastic dominance," in Stochastic Dominance, G. A. Whitmore and M. C. Findlay, eds., Lexington Books, 352-363.